

CEVO: Comprehensive EVent Ontology Enhancing Cognitive Annotation

Saeedeh Shekarpour Valerie Shalin
Krishnaprasad Thirunarayan Amit P. Sheth

January 23, 2017

Abstract

While the general analysis of named entities has received substantial research attention, the analysis of relations over named entities has not. In fact, a review of the literature on unstructured as well as structured data revealed a deficiency in research on the abstract conceptualization required to organize relations. We believe that such an abstract conceptualization can benefit various communities and applications such as natural language processing, information extraction, machine learning and ontology engineering. In this paper, we present CEVO (i.e., a Comprehensive EVent Ontology) built on Levin’s conceptual hierarchy of English verbs that categorizes verbs with the shared meaning and syntactic behavior. We present the fundamental concepts and requirements for this ontology. Furthermore, we present three use cases for demonstrating the benefits of this ontology on annotation tasks: 1) annotating relations in plain text, 2) annotating ontological properties and 3) linking textual relations to ontological properties.

1 Introduction

The size of structured as well as unstructured data on the Web is enormously growing. On one side, an important portion of the Web remains unstructured (i.e. textual data) which are being published from social network feeds, blogs, news, logs etc. On the other side, the size of structured data openly published is indeed considerable. So far, Linked Data has published more than 130 billion triples from over 9960 datasets¹. These large datasets result from heterogeneous, ad hoc ontologies and lexicons. While this heterogeneity provides flexibility, it complicates both reuse of ontologies and interlinking of datasets.

So far, Named entity recognizing as well as entity linking to the background knowledge base have received substantial research attention. But the analysis of relations over named entities has not. Although cognitive scientists (e.g., Doumas and Humel, 2005) suggest that relational content is key to reasoning, a review of the literature on unstructured as well as structured data revealed a deficiency in research on the abstract conceptualization required to organize relations. Observing deficiencies in (i)

¹observed on October 18th 2016 at <http://lodstats.aksw.org/>

relation extraction from text (ii) *contextual equivalencing of relations* and (iii) dealing with *diversity of ontologies* motivated us to investigate an abstract conceptualization on relations. Linguists such as Jackendoff and Talmy have a long history of concern with the conceptual structure of relations, our interest in text annotation requires a link between lexical items and a conceptual structure. We found an answer to our struggle in the organized lexicon and knowledge base assembled by the Stanford linguist Beth Levin in [7]. While Levin relies on Schank’s conceptual dependency theory [9, 10] to organize this knowledge base, the key here is the psychologically principled inventory of English verbs aligned with the knowledge base. Classes in this knowledge base identify sets of semantically coherent verbs with corresponding syntactic properties. For example, the *communication* class refers to verbs transferring a message/idea (i.e. shared meaning) such as *announce*, *say*, *mention*. In addition, they share the same syntactic behaviour such as NP1 VP NP2 or NP1 VP NP2 to NP3 (NP is noun phrase and VP is verb phrase). This conceptualization provides more than 230 classes for over 3000 English verbs. Using Levin’s work, we build an event ontology and lexicon called *CEVO*². CEVO is designed to recognize and equate relations from both textual data sources as well as knowledge bases. Such an abstract conceptualization benefits many applications such as natural language processing, information extraction, ontology engineering and machine learning.

Figure 1 represents the evolution of ontologies as well as vocabularies in terms of *level of abstraction*. Primarily, the early generation of vocabularies was created for annotating datasets (i.e. meta data) or describing a domain. The next generation of vocabularies are those used for interoperability issues, they are created from a higher abstraction level. The other generation of ontologies have cognitive applicability, thus they have the highest level of abstraction.

This paper is organized as follows. Section 2 presents the deficiencies that motivated us to develop the CEVO ontology. Section 3 discusses fundamental requirements for designing the CEVO ontology. The principles and considerations that Beth Levin has taken into account for categorizing English verbs are presented in Section 4. The main concepts of CEVO are introduced in Section 5. We discuss three use cases employing CEVO for annotation tasks in Section 6. Related work is presented in Section 8. We close with the conclusion and future work 9.

2 Problem Statement

The CEV-Ontology (CEVO) compensates for pervasive deficiencies in the abstract conceptualization of relations. In the following, we mention the three well-known deficiencies towards annotating relations or ontological properties.

- (I) *Relation Extraction*: Decades of research in the field of information extraction has resulted in developing tools that successfully recognize and annotate Named Entities (NE) and link them to entities available in background knowledge base (e.g., [5, 4, 8]). In contrast, recognizing, tagging and linking relationships among entities have received limited attention.

²CEVO namespace: <http://eventontology.org/>

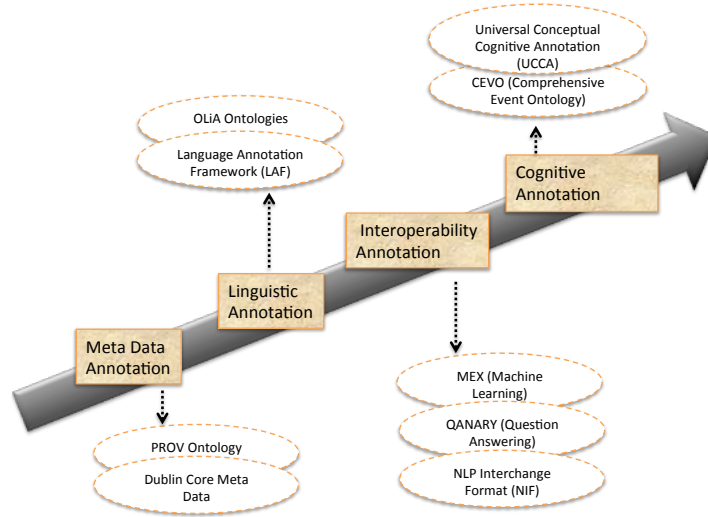


Figure 1: Evolution of Abstraction Level of Vocabularies and Ontologies

- (II) *Contextual Equivalence of Relations*: Relations embedded in plain text can be expressed in various ways either explicitly or implicitly. Explicit relations often appear as a verb phrase and implicit ones are usually hidden or embedded in other phrases, e.g., adjective phrase such as "taller than". Moreover, a single explicit relation can be expressed using several distinct verbs. E.g., consider the two sentences 'Jack visits Sara', and 'Jack consults Sara'. In both of these cases, the abstract event of *meeting* is expressed using two different verbs 'visit' and 'consult'. Nevertheless, these two verbs do not have a simple synonymous relationship recoverable by inference using lexicons such as WordNet. These verbs convey the same event in a specific context.
- (III) *Diversity in conceptualization*: Each ontology is created based upon a specific interpretation of a domain. This strategy yields heterogeneous ontologies due to different conceptualizations of an individual concept or property inherent in different users or communities. While this heterogeneity provides flexibility, it complicates reusability and introduces integration challenges. Linking the mention of either entity or relation from plain text to the corresponding background knowledge base is ontology-specific.

3 Requirements for CEVO

On the basis of these observations, we derive a series of requirements for a cognitive ontology for annotating relations on both structured as well as unstructured data. The coarse-grained requirements are listed below.

Requirement 1 (Relation Tagging on Textual Data) *Similar to tagging Named Entities in plain text, each mention of a relation must be recognized, normalized, and tagged. Thus, a tags set is required for distinguishing relations.*

Requirement 2 (Relation Linking) *Beyond recognizing and tagging mentions of relations in plain text, it is necessary to link textual relations to ontological properties. To do that, having an upper ontology which is used for annotating both ontological properties and textual relations is required.*

Requirement 3 (Integration and Alignment of Properties) *The variety of ways to conceptualize a domain results in different ontologies. Certainly, overlaps require alignment or integration. Thus, annotating ontologies based on an upper ontology which has a higher abstraction helps process of integrating and aligning ontologies.*

Requirement 4 (Reusability) *One of the main obstacles for the reuse of ontologies is the additional effort required for interpreting represented conceptualization. Providing annotations based on a cognitive conceptualization indeed boosts reusability.*

Requirement 5 (Simplicity) *Since we desire CEVO (a cognitive conceptualization over relations) to be widely adopted and reused, the captured cognitive conceptualization has to be as simple as possible to minimize integration and adoption efforts.*

To the best of our knowledge, there is no existing ontology on Linked Data that fulfills the above requirements. To this end, we present CEVO (a cognitive event ontology), built upon [7]. We contend that CEVO is fulfilled with an abstract conceptualization of relations. In the following section, we discuss principles behind this conceptual hierarchy.

4 Levin Conceptual Hierarchy

The entries of Levin’s lexical knowledge base are verb classes. For example, Figure 2 illustrates two distinct English verb classes (1) *transformation and creation* and (2) *change of the state* that both subsume several verbs . The members of each class (i.e. English verbs) have two characteristics: (i) semantically coherent and (ii) shared syntactic behavior. These characteristics are described as follows:

(i) *Shared meaning*: Each class of verbs shows a unique set of properties that shape the meaning of the member verbs. In fact, the conjunction of properties provides a distinctive meaning for each class. A single meaning property might be attributed to several classes depending upon context. Moreover, an individual verb might belong to multiple classes, creating a graph instead of a tree. For example, the class of *Creation and Transformation* (shown in Figure 2) refers to a class of verbs causing alternation. This class contains of both transitive verbs (an agent creates an entity) and intransitive verbs (describing transformation of an entity). A sub class of this class is *Build class* with the shared properties as (1) material/product alternation (2) total transformation alternation (3) unspecified object alternation (4) benefactive alternation (5) causative alternation (6) raw material subject alternation (7) sum of money subject alternation while another sub class namely *Grow class* only shares three properties as (1) material/product alternation (2) total transformation alternation (3) causative alternation. An English verb might belong to two distinct classes. For example, the verbs *cook* and *boil* belong to two distinct classes (a) creation and transformation and (b) change of state

(represented in in Figure 2). Thus, depending on the *context*, the appropriate class is distinguished.

(ii) *Shared syntactic behavior*: Meaning influences the syntactic behavior of a verb in terms of its expression and interpretation of arguments. Verbs with shared meaning exhibit similar syntactic behavior.

Change of State (cooking)	Creation And Transformation (Build)
<ul style="list-style-type: none"> •Roast •Cook •Grill •Broil •Crisp •Boil •Heat •Fry •... 	<ul style="list-style-type: none"> •Bake •Build •Cook •Develop •Grow •Weave •Shape •Boil •...

Figure 2: Two classes of Levin categorization with samples of member verbs: (i) Creation and transformation class and (ii) Change of the state class

5 CEVO: Comprehensive EEvent Ontology

In this section, we describe the main concepts introduced in CEVO in the schema as well as instance level. The namespace used in CEVO is <http://eventontology.org/>. This namespace is abbreviated as `cevo` in the following.

CEVO Schema: The top class of CEVO is the class of generic *Event* that is the superclass of all specific events. The generic Event class is formally defined as follows:

Definition 1 (Class of generic Event) *The generic ‘Event’ is an `owl:Class` and refers to ‘occurrence of anything’. It generally is the superclass of any specific type of event.*

```
cevo:#Event    a          owl:Class          .
cevo:#Event    rdfs:label  'generic event'      .
cevo:#Event    rdfs:comment 'something that happens' .
```

In CEVO, the Levin conceptual hierarchy is incorporated under the generic *Event* class. Figure 3 illustrates the first level of Levin’s hierarchy. In other words, any class provided for a set of English verbs revealing a specific event which is considered as an `owl:Class`. Formally as:

Definition 2 (Class of ‘X’ Event) *‘X’ Event is a subclass of the class Event. Conceptually it refers to a specific type of event that is associated with an English verb category sharing a common behavior or meaning.*

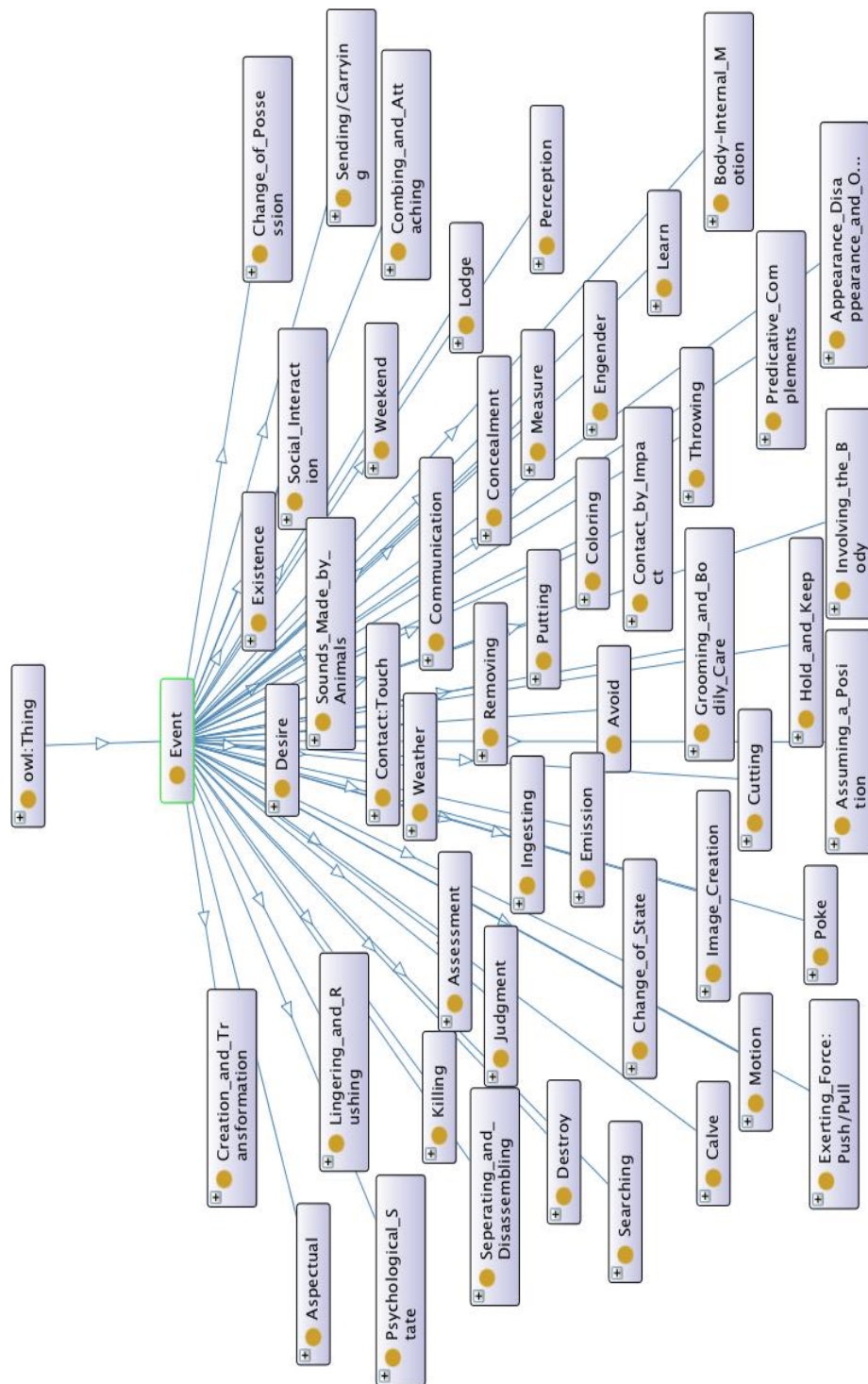


Figure 3: First level of event hierarchy

For instance, the class *communication event* given below is defined as a subclass of generic Event. This class refers to occurrence of any activity for communicating or transferring message/idea. Figure 4 represents the hierarchy on communication event in CEVO. Furthermore, this event is divided into eight sub events. For example, the sub event *complain* specifies the speakers' attitude or feeling towards what is said in addition to communicating activity.

```
cevo:#Communication    a                owl:Class                ;
                        rdfs:subClassOf    cevo:#Event                ;
                        rdfs:label         'communication'           ;
                        rdfs:comment       'communication and
                        transfer of idea'    .
```

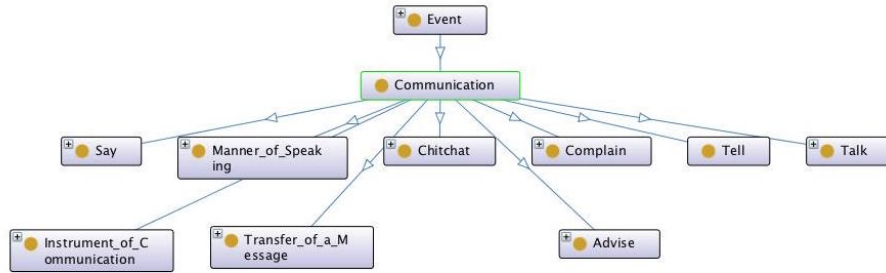


Figure 4: Hierarchy of Communication Event

The next main class is *cevo#MainVerb* that refers to words with part of speech as *verb*. This class is equivalent to the class of *main verb* of *OLiA*³ ontology [3, 2]. *OLiA*⁴ is an annotation model based on morphology.

```
cevo:#MainVerb    a                owl:Class                .
cevo:#MainVerb    owl:equivalentClass    OLiA:MainVerb    .
```

CEVO Instance Level (Verb Individuals) So far, we described the schema level classes; the next important step is to map each individual English verb to the corresponding event class. Thus, we instantiate each English verb at the instance level and type that primarily as *cevo:#MainVerb* and then map it to the associated event(s). In the following, two English verbs *say* and *cook* are instantiated. They are primarily typed as *cevo:#MainVerb* and furthermore, they are typed to their corresponding event classes respectively *Communication* and *Creation and Transformation* events. In fact, by specifying the type of a verb as *cevo:#MainVerb*, explicitly syntactic role of that verb in the English language is determined and by associating the relevant events to a verb, domain-specific semantic roles of that verb is determined. Please note that, each individual verb might be associated with several event classes. For instance,

³<http://nachhalt.sfb632.uni-potsdam.de/owl/olia.owl>

⁴<http://nachhalt.sfb632.uni-potsdam.de/owl/>

the verb *cook*, in addition to the event *Creation and Transformation*, is also associated with the event *Change of the State* (shown in Figure 5).

```

(a) cevo:#say    rdf:type    cevo:#MainVerb          .
(a) cevo:#say    rdf:type    cevo:#Communication      .
(b) cevo:#cook   rdf:type    cevo:#MainVerb          .
(b) cevo:#cook   rdf:type    cevo:#Creation_Transformation .
(b) cevo:#cook   rdf:type    cevo:#Change_of_the_state .
(b) cevo:#cook   rdf:type    cevo:#Cooking           .
(b) cevo:#cook   rdf:type    cevo:#Build             .

```

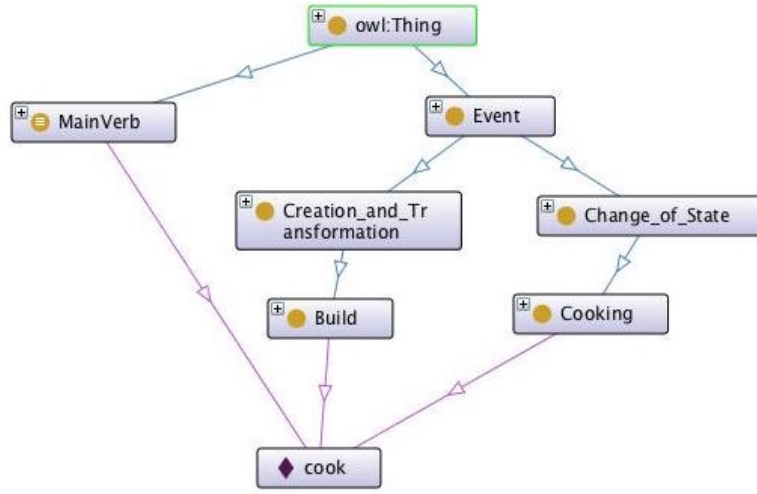


Figure 5: Associated classes to the verb *cook*.

On the other hand a group of verbs reveals occurring of an event. For example, Figure 6 represents all the verbs which may cause occurring the event of *Transferring a message* being a sub event of the Communication event.

6 Use Cases

In the following, we present three use cases showing employing CEVO for annotating textual relations and ontological relations.

6.1 Use Case 1: Annotating Relations in Text using CEVO

CEVO promotes annotating relations in plain text. Figure 7 shows two headline news on Twitter. The first tweet was published by BBC and the second one was published by New York Times. Tweet#1 is headed by the verb *announce* and the tweet#2 is

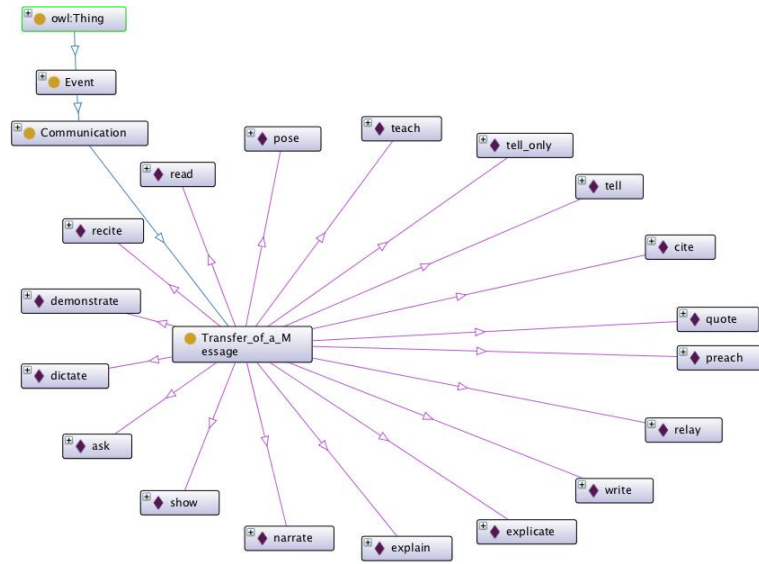


Figure 6: Verbs associated to the event of ‘transfer of a message’.

headed by the verb *say*. Both of these tweets have similar meaning in the sense that a *message is transferred*. Annotating these two tweets via CEVO leads us to obtain the same tag *communication* for both of these verbs, whereas the two verbs *announce* and *say* do not hold any lexical relations such as synonymy.

BBC Tweet#1 on 10/3/2016:

Obama and Justin Trudeau announce efforts to fight climate change.

CEVO:Communication

NYT Tweet#2 14/3/2016:

State elections were "difficult day," German Chancellor Angela Merkel says.

CEVO:Communication

Figure 7: Annotating relations expressed in headline news published on Twitter using CEVO. Both of these tweets obtained similar tags because, in both, an event of ‘transferring a message’ appears.

In the following, we annotate the two strings *announce* and *say* using nif⁵ vocabulary.

```
<exam:tweet#1#char=26, 33>      a      nif:String      ;
                                nif:beginIndex  26      ;
                                nif:endIndex    33      ;
```

⁵<http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core/nif-core.html>

```

                                nif:anchorOf      "announce"      ;
                                nif:oliaCategory   Olia:MainVerb .

<exam:tweet#2#char=71,74>    a                nif:String      ;
                                nif:beginIndex     71              ;
                                nif:endIndex       74              ;
                                nif:anchorOf      "says"          ;
                                nif:oliaCategory   Olia:MainVerb .

```

Both of these verbs `announce` and `say` obtain the relevant event type from `cevo` domain as follows:

```

<exam:tweet#1#char=26,33>    a                cevo:Communication .
<exam:tweet#2#char=71,74>    a                cevo:Communication .

```

6.2 Use Case 2: Annotating Properties of Ontologies

CEVO can be utilized for annotating properties of any ontology. One way of providing such an annotation is using the Web Annotation Data Model⁶ (WADM) which is a framework for expressing annotations. A WADM annotation has two elements (i) a target which indicates the resource being annotated and (ii) the body which indicates the description. Annotating properties of various ontologies according to CEVO addresses integration and alignment problems. Assume that we have the property `dbp:spouse`⁷ from DBpedia ontology that represents the relation of *marrying* that is semantically equivalent to the class `cevo:Amalgamate`. The annotation of this property is presented in Turtle syntax using WADM framework as follows:

```

example:annotation1    a                oa:Annotation      ;
                        oa:hasTarget     dbo:spouse        ;
                        oa:hasBody       cevo:Amalgamate    .

```

6.3 Use case 3: Linking Relations

CEVO facilitates linking occurrences of relations in plain text to ontological properties. We continue with the following example. On the 4th March 2016, BBC published this headline: `tweet#3:Rupert Murdoch and Jerry Hall marry`. The embedded relation in this part of text is `marry`. This relation is annotated as `cevo:Amalgamate` employing CEVO ontology. We show this annotation using nif vocabulary in the following.

```

<exam:tweet#3#char=31,35>    a                nif:String      ;
                                nif:beginIndex     31              ;
                                nif:endIndex       35              ;
                                nif:anchorOf      "marry"          ;

```

⁶ W3C Working Draft 15 October 2015, <http://www.w3.org/TR/annotation-model>

⁷We use the prefix `dbo` for DBpedia ontology.

```
nif:oliaCategory    Olia:MainVerb .

a                    cevo:Amalgamate .
```

Note that in our scenario, `example:headline1#marry` is the assigned URI for the verb `marry` on the mentioned headline. By taking into account the previous annotation for `dbo:spouse` property, now we are empowered to link the `marry` relation on the headline directly to the property `dbpedia:spouse` due to the similar tag of the CEVO ontology. This annotation is represented using WADM as follows:

```
example:annotation3  a          oa:Annotation          ;
                     oa:hasTarget example:headline1#marry ;
                     oa:hasBody   dbo:spouse           .
```

Thus, using SPARQL query, easily we can link a textual relation to the appropriate ontological relation based on their common CEVO annotations. For example the verb `marry` is linked to the property `dbpedia:spouse` as follows:

```
<exam:tweet#3#char=31,35>    itsrdf:taIdentRef

                                <\protect\vrule width0pt\protect\href{http://dbpedia.org/ontolo
```

7 Demo

8 Related Work

In this section, we review the definition of *event* given in different ontologies previously and then present several existing ontologies that facilitate the annotation task and interoperability among various components. LODE: An ontology for Linking Open Descriptions of Events⁸ defines a single generic concept of event as ‘Something that happened’, e.g., reported in a news article or with historical significance. This is a generic definition and does not specify various types of events necessary for subsequent inference. Schema.org⁹ introduces a similar generic concept of event¹⁰ that additionally considers temporal as well as location aspects and additionally provides a limited hierarchy. This hierarchy introduces types of events such as business events, sale events, and social events. Similarly, DBpedia ontology¹¹ defines a generic concept of event with a hierarchy which is broader, including lifecycle events (e.g. birth, death), natural events (e.g. earthquake, stormsurge) and societal events (e.g. concert, election). The interoperability among heterogenous datasets, schemas, various tools and dependent components is an important issue. So, recently, a number of linguistic ontologies

⁸<http://linkedevents.org/ontology/>

⁹<http://schema.org>

¹⁰<http://schema.org/Event>

¹¹<http://wiki.dbpedia.org/services-resources/ontology>

as well as annotation tools for related purposes have emerged. The Ontologies of Linguistic Annotation (OLiA)¹²[3, 2] provides an annotation tag sets using syntactical and morphological perspectives. OLiA covers over 110 OWL ontologies for over 34 tag sets in 69 different languages. Thus, all NLP tools can leverage OLiA's tag sets for annotating the output. Among the ontologies introduced for promoting interoperability between tools, services and components, we mention two recent ones:

1. NLP Interchange Format (NIF)¹³ [6] provides a vocabulary for interoperability between Natural Language Tools (NLP). NIF allows tools to exchange annotations for any part of text using the three layers: (1) *Structural layer*: This describes URI scheme for identifying any part of document, thus, making each piece of text dereferenceable. (2) *Conceptual layer*: NIF Core Ontology¹⁴ describes classes and properties for providing relations between tokens and documents. The core class is `nif:String` which points to any mentioned word in Unicode characters. (3) *Access layer*: NIF-aware applications publish their output using the NIF format via REST APIs.
2. QANARY¹⁵ [11, 1] is intended to facilitate interoperability among component of question answering systems. Currently, it defines a two layered ontology: *abstract layer* describing the generic functionalities of each components and *binding layer* enabling binding output of each component to the abstract layer.

To the best of our knowledge, CEVO is the first event ontology that provides a fine-grain abstract conceptualization of events. More importantly, CEVO connects this conceptualization to a large lexicon of English verbs. Potentially, this conceptualization will play a significant role in overcoming the existing challenges of (i) tagging relations, (ii) linking relations, and (iii) ontology alignment.

9 Conclusions and Future Directions

In this paper, we introduced an event ontology called *CEVO*¹⁶. This ontology relies on an abstract conceptualization over English verbs provided by Beth Levin in [7]. Such an abstract conceptualization largely obviates deficiencies in (i) relation extraction from text (ii) contextual equivalencing of relations and (iii) diversity of ontologies. This ontology presents more than 230 event classes for over 3000 English verbs as individuals. We plan to extend CEVO in the direction of integrating and interlinking to other existing ontologies, especially those that contain a conceptualization of events. Currently, we are applying CEVO to the domain of disaster response.

Acknowledgments We acknowledge partial support from the National Science Foundation (NSF) award: EAR 1520870: Hazards SEES: Social and Physical Sensing En-

¹²<http://www.acoli.informatik.uni-frankfurt.de/resources/olia/>

¹³<http://persistence.uni-leipzig.org/nlp2rdf/>

¹⁴<http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core/nif-core.html#>

¹⁵<https://github.com/WDAqua/QAontology>

¹⁶Homepage of the project: <http://wiki.knoesis.org/index.php/CEVO>

abled Decision Support for Disaster Management and Response. Any opinions, findings, and conclusions/recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

References

- [1] Andreas Both, Jens Lehmann, Sören Auer, and Martin Brümmer. Integrating NLP using linked data. In *The Semantic Web - ISWC 2013 - 11th International Extended Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part II*, pages 98–113, 2013.
- [2] Christian Chiarcos. Ontologies of linguistic annotation: Survey and perspectives. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC’12)*, Istanbul, Turkey, may 2012. European Language Resources Association (ELRA).
- [3] Christian Chiarcos and Maria Sukhareva. Olia - ontologies of linguistic annotation. *Semantic Web*, 6(4):379–386, 2015.
- [4] Paolo Ferragina and Ugo Scaiella. Fast and accurate annotation of short texts with wikipedia pages. *IEEE Software*, 29(1):70–75, 2012.
- [5] Jenny Rose Finkel, Trond Grenager, and Christopher Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, ACL ’05*, 2005.
- [6] Sebastian Hellmann, Jens Lehmann, Sören Auer, and Martin Brümmer. Integrating NLP using linked data. In *The Semantic Web - ISWC 2013 - 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part II*, pages 98–113, 2013.
- [7] Beth Levin. *English Verb Classes and Alternations:~A Preliminary Investigation*. University of Chicago Press, 1993.
- [8] Pablo N. Mendes, Max Jakob, Andrés García-Silva, and Christian Bizer. Dbpedia spotlight: shedding light on the web of documents. In *Proceedings the 7th International Conference on Semantic Systems, I-SEMANTICS 2011, Graz, Austria, September 7-9, 2011*, pages 1–8, 2011.
- [9] R. C. Schank. Conceptual dependency: A theory of natural language processing. 3, 1972.
- [10] R.C. Schank. *Conceptual Information Processing*. North-Holland, Amsterdam, 1975.
- [11] Kuldeep Singh, Andreas Both, Dennis Diefenbach, and Saeedeh Shekarpour. Towards a message-driven vocabulary for promoting the interoperability of question answering systems. In *Tenth IEEE International Conference on Semantic Computing, ICSC 2016, Laguna Hills, CA, USA, February 4-6, 2016*, pages 386–389, 2016.